Official

--The feature extraction module 202 performs edge-detection, signal conditioning and feature extraction. According to one embodiment, speech edge detection is accomplished using noise estimation and energy detection based on the 0^{th} Cepstral coefficient and zero-crossing statistics. Feature extraction and signal conditioning consist of extracting Mel-frequency cepstral coefficients (MFCC), delta information and acceleration information. It is a 38 dimensional feature vector based on 12.8 ms sample buffers overlapped by 50%. Such feature extraction modules 202 and functionality are well understood in the art, and one skilled in the art may implement the feature extraction module in a variety of ways. Thus, the output of the feature extraction module 202 is a sequence of feature vectors.--

N.E. Please replace the paragraph beginning at page 26, line 19, with the following rewritten paragraph:

--In some embodiments, a caching scheme is used for the lexicons stored in memory on the remote unit, e.g., by the N-gram grammar module 218. A stated above, a lexicon is a

dictionary consisting of words and their pronunciation entries. These pronunciations may be implemented as either phonetic spellings that refer to phonetic models, or to whole-word models. A given word entry may contain alternate pronunciation entries, most of which are seldom used by any single speaker. This redundancy is echoed at each part-of-speech abstraction, creating even more entries that are never utilized by a given speaker. This implies that if lexicon entries are sorted by their frequency of usage, there is a great chance that the words in an utterance can be found among the top *n* lexicon entries. As such, the cache is divided into different levels divided by frequency of use. For example, frequently used lexicon entries will be stored within the top level of the cache. A caching scheme may be devised in which the top 10% of the cache is used 90% of the time, for example. Thus, according to an embodiment, a multi-pass search is performed where the most likely entries are considered in the first pass. If the garbage score from this pass is high enough to believe that the words actually spoken were contained in the set of most likely spellings, the speech decoder 216 reports the results to the calling function.

If this score is low, the system falls back to considering a wider range of spellings. If the score from the first pass is high, but not high enough in order to be able to make a decision whether the correct spellings, for the elements of the utterance, were contained in the set of most likely spellings, this is also reported back to the calling function, which might prompt the user for clarification. If a lexicon spelling for a given part-of-speech is never used while some of its alternative spellings are frequently used, that spelling is put in a "trash can" and will never be considered for that user. As such, rarely used spellings are not considered and the chance of confusing similar-sounding utterances with one of those spellings is reduced and the recognition accuracy is therefore increased. Further, the caching scheme allows the system to consider less data and hence provides a great speed improvement.--

N.R. Please replace the paragraph beginning at page 35, line 1, with the following rewritten paragraph:

--Furthermore, the NLICS 102 may download command signals for the device abstraction module of the remote unit 104. For example, a user would like to operate an older VCR that has an IR remote control manufactured by a different maker than the NLICS. The base unit 106 simply downloads the commands that are stored for any number of devices. These commands are then stored in the device abstraction module. Also, the NLICS can submit feature vector data and labels associated with high-confidence utterances to the collaborative corpus. This data can then be incorporated with other data and used to train improved models that are subsequently redistributed. This approach can also be used to incorporate new words into the collaborative corpus by submitting the feature vector data and its label, which may subsequently be combined with other data and phonetically transcribed using the forward-backward algorithm. This entry may then be added to the lexicon and redistributed.--

12. Please replace the paragraph beginning at page 41, line 3, with the following rewritten paragraph: